Degree Distribution for Duplication-Divergence Graphs: Large Deviations*

Alan Frieze[1][0000−0002−8481−5615], Krzysztof Turowski[2][0000−0003−3871−1038], and Wojciech Szpankowski[3][0000−0001−9062−0067]

[1] Department of Mathematical Sciences,
Carnegie Mellon University, Pittsburgh, PA, USA
alan@random.math.cmu.edu
[2] Department of Theoretical Computer Science, Jagiellonian University, Poland
krzysztof.szymon.turowski@gmail.com
[3] Center for Science of Information, Department of Computer Science,
Purdue University, West Lafayette, IN, USA
spa@cs.purdue.edu

Abstract. We present a rigorous and precise analysis of the degree distribution in a dynamic graph model introduced by Solé et al. in which nodes are added according to a duplication-divergence mechanism, i.e. by iteratively copying a node and then randomly inserting and deleting some edges for a copied node. This graph model finds many applications since it well captures the growth of some real-world processes e.g. biological or social networks. However, there are only a handful of rigorous results concerning this model. In this paper we present rigorous results concerning the degree distribution.
We focus on two related problems: the expected value and large deviation for the *degree of a fixed node* through the evolution of the graph and the expected value and large deviation of the *average degree* in the graph. We present exact and asymptotic results showing that both quantities may decrease or increase over time depending on the model parameters. Our findings are a step towards a better understanding of the overall graph behaviors, especially, degree distribution, symmetry, and compression, important open problems in this area.

Keywords: Dynamic graphs · Duplication-divergence graphs · Degree distribution · Large deviation.

1 Introduction

It is widely accepted that we live in the age of data deluge. On a daily basis we observe the increasing availability of data collected and stored in various forms,

* This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grant CCF-1524312, and National Science Center, Poland, Grant 2018/31/B/ST6/01294. This work was also supported by NSF Grant DMS1661063.

as sequences, expressions, interactions or structures. A large part of this data is given in a complex form which also conveys a "shape" of the structure, such as network data. As examples we have various biological networks, social networks, and Web graphs.

Given a representation of these networks as graphs, there arises a natural question: what are the rules governing the growth and evolution of such networks? In fact, finding such rules should enable us to model real networks arising in many diverse applications. For example, there is experimental evidence [17] that the evolution of some biological networks is driven by the duplication mechanism, in which new nodes appear as copies of some already existing ones in the network. This is supplemented by a certain amount of divergence due to random mutations that leads to some differences between patterns of interaction for the source and the duplicate elements.

Fundamental questions arise about the structural properties of these networks. For example, Faloutsos, Faloutsos and Faloutsos [5] brought to the front the issue of "scale-free" power law behavior. First, there is the question as to whether or not the degree distributions of the real-world networks do indeed have a tail close to a power law. Second and most importantl, one needs to verify whether the underlying random graph models may indeed generate graphs that exhibit the desired behavior e.g. in expectation. This is directly related to the broader question of the degree distributions in graphs that may be generated from these random models. We would expect that a good model for real-world networks generates graphs that typically are not much different in terms of the number of vertices with given degrees to what is seen in practice. However, to answer this question we first need a good theoretical understanding of the degree distribution of our models and this is the subject of this paper.

Another important problem in this area is the question of *symmetry.* It may be formulated as follows: given a probability distribution over graphs of size n, what is the distribution of $\log |\mathrm{Aut}(G)|$, where G is a random graph drawn from this distribution and $\mathrm{Aut}(G)$ is its automorphism group (i.e., permutation preserving adjacency). Clearly, it is related to the degree distribution problem since for example the number of small symmetrical structures like cherries and diamonds (vertices of degree 1 and 2, respectively, having the same neighborhood) is a lower bound on the number of automorphisms for any graph. Interestingly enough, many real-world networks such as protein-protein and social networks, exhibit a lot of symmetry as shown in Table 1.

It turns out that the most popular random graph models do not exhibit much symmetry. For example, it was proved in [2] that the Erdős-Renyi random graph model generates asymmetric graphs (for not too small and too large edge probability), that is, $\log |\mathrm{Aut}(G)| = 0$, with high probability. In a similar vein, it was proved for the preferential attachment model, also known as the Barabási-Albert model, that for $m \geq 3$ (which is necessary if we want to obtain sufficiently dense graphs that resemble real-world networks) is asymmetric with high probability [9].

Network	Nodes	Edges	$\log\|\mathrm{Aut}(G)\|$
Baker's yeast protein-protein interactions	6,152	531,400	546
Fission yeast protein-protein interactions	4,177	58,084	675
Mouse protein-protein interactions	6,849	18,380	305
Human protein-protein interactions	17,295	296,637	3026
ArXiv high energy physics citations	7,464	116,268	13
Simple English Wikipedia hyperlinks	10,000	169,894	1019
CollegeMsg online messages	1,899	59,835	232

Table 1: Symmetries of the real-world networks [13, 16].

Therefore, in order to study and understand the behavior of real-world networks we need to look at dynamic graph models that naturally generate internal graph symmetries. As was mentioned before, one promising route is to investigate models that evolve according to the duplication and mutation rules. So let us consider the most popular duplication-divergence model introduced by Solé et al. [12], referred to below as $\mathtt{DD}(t, p, r)$. It is defined as follows: starting from a given graph on t_0 vertices (labeled from 1 to t_0) we add subsequent vertices labeled $t_0 + 1$, $t_0 + 2$, ..., t as copies of some existing vertices in the graph and then we introduce divergence by adding and removing some edges connected to the new vertex independently at random. Finally, we remove the labels and return the structure, i.e. the unlabeled graph.

It has been shown that for a certain set of parameters, graphs generated according to the duplication-divergence mechanism fit very well empirically with the structure of some real-world networks (e.g., protein-protein and citation networks) in terms of the degree distribution [3] and small subgraph (graphlets) counts [11]. However, at the moment there do not exist any rigorous general results regarding symmetries and hence degree distribution for such graphs. Experimentally, when generating multiple graphs according to this model with different parameters, we observe the pattern presented in Figure 1: There is a set of parameters (i.e., p and r) for which the generated graphs are highly symmetric with a large automorphisms group. It was shown by Sreedharan et al. [13] that possible values of the parameters for real-world networks lie in the blue-violet area, indicating a lot of symmetry. All these remarks suggest that there is a certain merit to study a possible link between the duplication-divergence model and certain types of real-world networks. To accomplish this we need to study the average and large deviation of their degree sequence, which is the main topic of this conference paper.

There exist only a handful previous rigorous results on the $\mathtt{DD}(t, p, r)$ model. In view of these, it is imperative that we understand the degree distribution for the duplication-divergence networks. Turowski et al. showed in [15] that for the special case of $p = 1$, $r = 0$ the expected logarithm of the number of automorphisms for graphs on t vertices is asymptotically $\Theta(t \log t)$, which indicates a lot of symmetry. This allows to construct asymptotically optimal compression algorithms for such graphs. However, the proposed approach used certain properties for this particular set of parameters that does not generalize to other sest of parametesr.

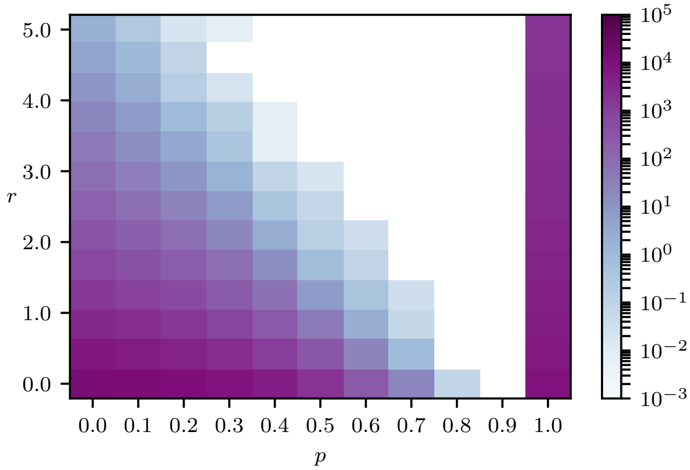

Fig. 1: Symmetry of graphs ($\log|\mathrm{Aut}(G)|$) generated by the $\mathtt{DD}(t, p, r)$ model.

For $r = 0$ and $p < 1$, it was recently proved by Hermann and Pfaffelhuber in [6] that depending on the value of p either there exists a limiting distribution of degree frequencies with almost all vertices isolated or there is no limiting distribution as $t \to \infty$. Moreover, it is shown in [8] that the number of vertices of degree one is $\Omega(\log t)$ but again the precise rate of growth of the number of vertices with any fixed degree $k > 0$ is currently unknown. Recently, also for $r = 0$, Jordan [7] showed that the *non-trivial connected component* has a degree distribution which conforms to a power-law behavior in size, but only for $p < e^{-1}$. In this case the exponent is equal to γ which is the solution of $3 = \gamma + p^{\gamma-2}$.

In this paper we study the degree distribution from a different perspective. In particular, we present results concerning the degree of a given vertex s at time t (denoted by $\deg_t(s)$) and the average degree in the graph (denoted by $D(G_t)$). We show that the asymptotic values of the means $\mathbb{E}[\deg_t(s)]$ and $\mathbb{E}[D(G_t)]$ as $t \to \infty$ exhibit phase transitions over the parameter space as a function of p and r. We then present some results for the tails of the degree distribution for $D(G_t)$ and $\deg_t(s)$ for $s = O(1)$. It turns out that the deviation by a polylogarithmic factor under or over the respective means is sufficient to obtain a polynomial tail, that is to find an $O(t^{-A})$ tail probability. In this way we have proved that the distribution of $D(G_t)$ and $\deg_t(s)$ are in some sense concentrated around their means.

2 Main results

In this section we first define formally the $\mathtt{DD}(t, p, r)$ model. Then, we present our main results, first concerning the expected values of the degree distribution, and then large deviations of their distributions. Here we provide only the main lines of reasoning, with the full proofs moved to the respective appendices.

We use standard graph notation, e.g. from [4]: $V(G)$ denotes the set of vertices of graph G, $\mathcal{N}_G(u)$ – the set of neighbors of vertex u in G, $\deg_G(u) = |\mathcal{N}_G(u)|$ – the degree of u in G. For brevity we use the abbreviations for G_t, e.g. $\deg_t(u)$ instead of $\deg_{G_t}(u)$. All graphs are simple. Let us also introduce the *average degree* $D(G_t)$ of G as

$$D(G) = \frac{1}{|V(G)|} \sum_{u \in V(G)} \deg_G(u).$$

It is worth noting that it is also known in the literature as the first moment of the degree distribution.

We formally define the model $\mathtt{DD}(t, p, r)$ as follows: let $0 \le p \le 1$ and $0 \le r \le t_0$ be the parameters of the model. Let also G_{t_0} be a graph on t_0 vertices, with $V(G_{t_0}) = \{1, \dots, t_0\}$. Now, for every $t = t_0, t_0 + 1, \dots$ we create G_{t+1} from G_t according to the following rules:

1. add a new vertex $t + 1$ to the graph,
2. pick vertex u from $V(G_t) = \{1, \dots, t\}$ uniformly at random – and denote u as *parent*$(t + 1)$,
3. for every vertex $i \in V(G_t)$:
   (a) if $i \in \mathcal{N}_t(\textit{parent}(t + 1))$, then add an edge between i and $t + 1$ with probability p,
   (b) if $i \notin \mathcal{N}_t(\textit{parent}(t + 1))$, then add an edge between i and $t + 1$ with probability $\frac{r}{t}$.

2.1 Average degree in the graph

We start with the average degree in the graph $D(G_t)$. First, we find the following recurrence for the average degree of G_{t+1}:

$$\begin{aligned}
\mathbb{E}[D(G_{t+1}) \mid G_t] &= \frac{1}{t+1} \mathbb{E}\left[\sum_{i=1}^{t+1} \deg_{t+1}(i) \;\middle|\; G_t\right] \\
&= \frac{1}{t+1} \mathbb{E}\left[\sum_{i=1}^{t} \deg_t(i) + 2\deg_{t+1}(t+1) \;\middle|\; G_t\right] \\
&= \frac{1}{t+1} \left(\sum_{i=1}^{t} \deg_t(i) + 2\mathbb{E}\left[\deg_{t+1}(t+1) \mid G_t\right]\right) \\
&= \frac{1}{t+1} \left(tD(G_t) + 2\mathbb{E}[\deg_{t+1}(t+1) \mid G_t]\right).
\end{aligned}$$

Next, we find the following relationship between the expected average degree $\mathbb{E}[D(G_t)]$ and the expected degree of the new vertex $\mathbb{E}[\deg_{t+1}(t + 1)]$ (see proof in Appendix A):

Lemma 1. *For any $t \ge t_0$ it holds that*

$$\mathbb{E}[\deg_{t+1}(t+1)] = \left(p - \frac{r}{t}\right) \mathbb{E}[D(G_t)] + r.$$

It is quite intuitive that the expected degree of a new vertex behaves as if we would choose a vertex with the average degree $\mathbb{E}[D(G_t)]$ as its parent, and then copy a p fraction of its edges, also adding r more edges (in expectation) to the other vertices in the graph.

From this lemma we find

$$\mathbb{E}[D(G_{t+1}) \mid G_t] = D(G_t)\left(1 + \frac{2p-1}{t+1} - \frac{2r}{t(t+1)}\right) + \frac{2r}{t+1}. \tag{1}$$

This recurrence falls under a general recurrence of the form

$$\mathbb{E}[f(G_{t+1}) \mid G_t] = f(G_t)g_1(t) + g_2(t) \tag{2}$$

where g_1 and g_2 are given functions, dependent on p and r. We will solve it exactly and asymptotically in the sequel. This allows us to find an asymptotic expression for the average degree.

In the sequel we present a series of lemmas that will be used to obtain the asymptotics of $\mathbb{E}[f(G_t)]$. These lemmas are based on martingale theory and they use various asymptotic properties of Euler gamma function. For space reasons, proofs of Lemmas 2 and 5 are moved to Appendices B and C.

Lemma 2. *Let $(G_n)_{n=n_0}^{\infty}$ be a Markov process for which $\mathbb{E}f(G_{n_0}) > 0$ and Equation (2) holds with $g_1(n) > 0$, $g_2(n) \geq 0$ for all $n = n_0, n_0+1, \ldots$. Then for all $n \geq n_0$*

$$\mathbb{E}f(G_n) = f(G_{n_0})\prod_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j)\prod_{k=j+1}^{n-1} g_1(k).$$

The above lemma shows that the solutions of recurrences of type Equation (2) contain products and sum of products of g_1 and g_2. The next lemmas show how to handle such products. First, since in our case $g_1(n)$ and $g_2(n)$ are of form $\frac{W_1(n)}{W_2(n)}$ for certain polynomials $W_1(n)$, $W_2(n)$, we turn the products of polynomials into the products of Euler gamma functions.

Lemma 3. *Let $W_1(k)$, $W_2(k)$ be polynomials of degree d with respective (not necessarily distinct) roots a_i, b_i $(i = 1, \ldots, d)$, that is, $W_1(k) = \prod_{i=1}^{d}(k - a_i)$ and $W_2(k) = \prod_{j=1}^{d}(k - b_j)$. Then*

$$\prod_{k=n_0}^{n-1} \frac{W_1(k)}{W_2(k)} = \prod_{i=1}^{d} \frac{\Gamma(n-a_i)}{\Gamma(n-b_i)} \frac{\Gamma(n_0-b_i)}{\Gamma(n_0-a_i)}.$$

Proof. We have

$$\prod_{k=n_0}^{n-1} \frac{W_1(k)}{W_2(k)} = \prod_{k=n_0}^{n-1}\prod_{i=1}^{d} \frac{k-a_i}{k-b_i} = \prod_{i=1}^{d}\prod_{k=n_0}^{n-1} \frac{k-a_i}{k-b_i} = \prod_{i=1}^{d} \frac{\Gamma(n-a_i)}{\Gamma(n-b_i)} \frac{\Gamma(n_0-b_i)}{\Gamma(n_0-a_i)}$$

which completes the proof.

Next, for the sake of completeness we present a well-known asymptotic formula for the Euler gamma function, which help us to deal with $\prod_{k=n_0}^{n-1} g_1(k)$:

Lemma 4 (Abramowitz, Stegun [1]). *For any $a, b \in \mathbb{R}$ if $n \to \infty$, then* $\frac{\Gamma(n+a)}{\Gamma(n+b)} = \Theta(n^{a-b})$.

Finally, we deal with sum of products $\sum_{j=n_0}^{n-1} g_2(j) \prod_{k=j+1}^{n-1} g_1(k)$. In terms of Euler gamma functions, via Lemma 3, we are interested in the asymptotics of the following formulas

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k} \Gamma(j+a_i)}{\prod_{i=1}^{k} \Gamma(j+b_i)}$$

with $a = \sum_{i=1}^{k} a_i$, $b = \sum_{i=1}^{k} b_i$.

Lemma 5. *Let $a_i, b_i \in \mathbb{R}$ $(k \in \mathbb{N})$ with $a = \sum_{i=1}^{k} a_i$, $b = \sum_{i=1}^{k} b_i$. Then it holds asymptotically for $n \to \infty$ that*

$$\sum_{j=n_0}^{n} \frac{\prod_{i=1}^{k} \Gamma(j+a_i)}{\prod_{i=1}^{k} \Gamma(j+b_i)} = \begin{cases} \Theta\left(n^{a-b+1}\right) & \text{if } a+1 > b, \\ \Theta\left(\log n\right) & \text{if } a+1 = b, \\ \Theta\left(1\right) & \text{if } a+1 < b. \end{cases}$$

With this background information, we are now in the position to solve recurrence (1) and present exact and asymptotic results for the average degree. From Lemma 2 with $g_1(t) = 1 + \frac{p}{t} - \frac{r}{t^2}$ and $g_2(t) = \frac{r}{t}$ we get that

$$\begin{aligned} \mathbb{E}[D(G_t)] &= D(G_{t_0}) \prod_{k=t_0}^{t-1} \left(1 + \frac{2p-1}{k+1} - \frac{2r}{k(k+1)}\right) \\ &\quad + \sum_{j=t_0}^{t-1} \frac{2r}{j+1} \prod_{k=j+1}^{t-1} \left(1 + \frac{2p-1}{k+1} - \frac{2r}{k(k+1)}\right). \end{aligned}$$

By applying Lemma 3 we replace the products above by the products of Euler gamma functions and obtain

$$\begin{aligned} \mathbb{E}[D(G_t)] &= D(G_{t_0}) \frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(t)\Gamma(t+1)} \frac{\Gamma(t+c_3)\Gamma(t+c_4)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} \\ &\quad + \sum_{j=t_0}^{t-1} \frac{2r\Gamma(j+1)}{\Gamma(j+2)} \frac{\Gamma(j+c_3)\Gamma(j+c_4)}{\Gamma(t_0+c_3)\Gamma(t_0+c_4)} \frac{\Gamma(t_0)\Gamma(t_0+1)}{\Gamma(j)\Gamma(j+1)} \end{aligned}$$

for c_3, c_4 – the (possibly complex) solutions of the equation $k^2 + 2pk - 2r = 0$.

Next, we may use Lemmas 4 and 5 respectively for the first and second part of the equation above. Note that when $r = 0$, the second part vanishes. Using previous lemmas we can derive asymptotics for the average degree which we present next.

Theorem 1. *For $G_t \sim \mathtt{DD}(t,p,r)$ asymptotically as $t \to \infty$ we have*

$$\mathbb{E}[D(G_t)] = \begin{cases} \Theta(1) & \text{if } p < \frac{1}{2} \text{ and } r > 0, \\ \Theta(\log t) & \text{if } p = \frac{1}{2} \text{ and } r > 0, \\ \Theta(t^{2p-1}) & \text{if } p > \frac{1}{2} \text{ or } r = 0. \end{cases}$$

The asymptotic behavior of $\mathbb{E}[D(G_t)]$ has a threefold characteristic: when $p < \frac{1}{2}$ and $r > 0$, the majority of the edges are not created by copying them from parents, but actually by attaching them according to the value of r. For $p = \frac{1}{2}$ and $r > 0$ we note the curious situation of a phase transition (still with non-copied edges dominating), and only if $p > \frac{1}{2}$ or $r = 0$ do the edges copied from the parents asymptotically contribute the major share of the edges.

The next question regarding the average degree $D(G_t)$ is how much it deviates from the expected value $\mathbb{E}[D(G_t)]$, in probability. It turns out that $D(G_t)$ is concentrated around $\mathbb{E}[D(G_t)]$ in such a way that with probability $1 - O(t^{-A})$ it falls within a polylogarithmic ratio from the mean. We observe that unlike the large deviations for say preferential attachment graphs, in the duplication-divergence model we need to consider three cases reflecting the different behavior of $\mathbb{E}[D(G_t)]$ for $p < 1/2$, $p = 1/2$ and $p > 1/2$.

Theorem 2. *Asymptotically for $G_t \sim \mathtt{DD}(t,p,r)$ it holds that*

$$\begin{aligned} &\Pr[D(G_t) \geq A\,C\, \log^2(t)] = O(t^{-A}) && \text{for } p < \frac{1}{2}, \\ &\Pr[D(G_t) \geq A\,C\, \log^3(t)] = O(t^{-A}) && \text{for } p = \frac{1}{2}, \\ &\Pr[D(G_t) \geq A\,C\, t^{2p-1} \log^2(t)] = O(t^{-A}) && \text{for } p > \frac{1}{2}. \end{aligned}$$

for some fixed constant $C > 0$ and any $A > 0$.

The complete proof of this theorem is presented in Appendix D. Here we outline the main steps. We begin by carefully bounding from above the moment generating function $\mathbb{E}[\exp(\lambda_t D(G_t))\,|G_t]$. This way, we are able to show

$$\mathbb{E}\left[\exp(\lambda_{t+1} D(G_{t+1})) \;\middle|\; G_t\right] \leq \exp\left(\lambda_{t+1} D(G_t) h_1(t) + \frac{\lambda_{t+1}}{t+1} h_2(t)\right)$$

for certain explicit functions h_1 and h_2. By defining $\lambda_t = \lambda_{t+1} h_1(t)$ we finally arrive at

$$\mathbb{E}[\exp(\lambda_t D(G_t))] \leq \exp(\lambda_{t_0} D(G_{t_0})) \left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1} + C_1}$$

for some $\varepsilon_t = 1/\ln(t/t_0)$ and constant C_1 (see (4) in Appendix D). After choosing a suitable sequence sequence[4] $(\lambda_i)_{i=t_0}^t$ and applying Chernoff's inequality we find

[4] In Appendix D we choose $\lambda_t = \varepsilon_t \left(\frac{t}{t_0}\right)^{-(2p-1)(1+O(\varepsilon_t))}$ so that $\lambda_{t_0} \leq \varepsilon_t$.

the large deviation bound. In all three cases we need an extra $\log^2(t)$ factor over the mean, one logarithm coming from the choice of ε_t (which decays like $\frac{1}{\ln t}$), and one from our requirement to get $O(t^{-A}) = O(\exp(-A\ln(t)))$ tail.

The left tail behavior is similar, however the proof presented in Appendix E is slightly more complicated, as discussed briefly below.

Theorem 3. *For $G_t \sim \mathtt{DD}(t, p, r)$ with $p > \frac{1}{2}$ asymptotically it holds that*

$$\Pr\left[D(G_t) \leq \frac{C}{A} t^{2p-1} \log^{-3-\varepsilon}(t)\right] = O(t^{-A}).$$

for some fixed constant $C > 0$ and any $\varepsilon, A > 0$.

Note that since $\mathbb{E}[D(G_t)] = O(\log t)$ for $p \leq \frac{1}{2}$, bounds of the above form are trivial in this range of p and therefore not interesting, since all smaller values are in polylogarithmic distance to the mean.

The whole proof, as presented in Appendix E, may be sketched as following: first, we find a bound

$$\mathbb{E}[\exp\left(\lambda(D(G_{t+1}) - D(G_t))\right) | G_t, \neg\mathcal{B}_t]$$

for a certain event $\mathcal{B}_t$ that allows us to bound the right tail for the variable $D(G_{t+1}) - D(G_t)$. Then, we use an auxiliary variable $Y_k = D(G_{(k+1)t}) - D(G_{kt})$ for which we know both the value of $\mathbb{E}[Y_k]$ and that the right tail of Y_k is small – in particular, it is $O(t^{-A})$ when we are only $\log^2(t)$ times over the mean. Therefore it may be shown that also the left tail of Y_k cannot be large. Finally, we use the result for Y_k to obtain a bound for $D(G_t)$.

2.2 Degree of a given vertex s

We focus now on the expected value of $\deg_t(s)$, that is, the degree of vertex s at time t. We start with a recurrence relation for $\mathbb{E}[\deg_t(s)]$. Observe that for any $t \geq s$ we know that vertex s may be connected to vertex $t+1$ in one of the following two cases:

- either $s \in \mathcal{N}_t(\mathit{parent}(t+1))$ (which holds with probability $\frac{\deg_t(s)}{t}$) and we add an edge between s and $t+1$ (with probability p),
- or $s \notin \mathcal{N}_t(\mathit{parent}(t+1))$ (with probability $\frac{t-\deg_t(s)}{t}$) and we an add edge between s and $t+1$ (with probability $\frac{r}{t}$).

From the model description we directly obtain the following recurrence for $\mathbb{E}[\deg_t(s)]$:

$$\begin{aligned}\mathbb{E}[\deg_{t+1}(s) \mid G_t] = &\left(\frac{\deg_t(s)}{t} p + \frac{t - \deg_t(s)}{t} \frac{r}{t}\right)(\deg_t(s) + 1) \\ &+ \left(\frac{\deg_t(s)}{t}(1-p) + \frac{t-\deg_t(s)}{t}\left(1 - \frac{r}{t}\right)\right)\deg_t(s)\end{aligned}$$

$$= \deg_t(s)\left(1+\frac{p}{t}-\frac{r}{t^2}\right)+\frac{r}{t}.$$

Again this recurrence falls under Equation (2), so we may proceed in the same fashion as with $\mathbb{E}[D(G_t)]$. First, we apply Lemma 2 with $g_1(t) = 1+\frac{p}{t}-\frac{r}{t^2}$ and $g_2(t) = \frac{r}{t}$ to obtain the equation for the exact behavior of the degree of a given node s at time t:

$$\mathbb{E}[\deg_t(s)] = \mathbb{E}[\deg_s(s)]\prod_{k=s}^{t-1}\left(1+\frac{p}{k}-\frac{r}{k^2}\right)+\sum_{j=s}^{t-1}\frac{r}{j}\prod_{k=j+1}^{t-1}\left(1+\frac{p}{k}-\frac{r}{k^2}\right).$$

Again we substitute the simple products by the products of Euler gamma functions using Lemma 3

$$\begin{aligned}\mathbb{E}[\deg_t(s)] &= \mathbb{E}[\deg_s(s)]\frac{\Gamma(s)^2}{\Gamma(t)^2}\frac{\Gamma(t+c_1)\Gamma(t+c_2)}{\Gamma(s+c_1)\Gamma(s+c_2)}\\ &\quad+\sum_{j=s}^{t-1}\frac{r\Gamma(j)}{\Gamma(j+1)}\frac{\Gamma(j+c_1)\Gamma(j+c_2)}{\Gamma(s+c_1)\Gamma(s+c_2)}\frac{\Gamma(s)^2}{\Gamma(j)^2}\end{aligned}$$

for c_1, c_2 – the (possibly complex) solutions of the equation $k^2+pk-r=0$.

We are finally in a position to state the asymptotic expressions for $\mathbb{E}[\deg_t(s)]$, using Lemmas 4 and 5.

Theorem 4. *For $s = O(1)$ it holds asymptotically as $t \to \infty$ that*

$$\mathbb{E}[\deg_t(s)] = \begin{cases}\Theta(\log t) & \text{if } p = 0 \text{ and } r > 0\\ \Theta(t^p) & \text{otherwise.}\end{cases}$$

Here we observe only two regimes. In the first, for the case when $p = 0$, when edges are added only due to the parameter r, we have logarithmic growth of $\mathbb{E}[\deg_t(s)]$. In the second one, edges attached to s accumulate mostly by choosing vertices adjacent to s as parents of the new vertices, and therefore the expected degree of s grows proportional to t^p.

If we assume that $s \to \infty$, that is, we consider asymptotics with respect to both s and t, we may combine Lemma 1 with Theorem 1 to obtain

Theorem 5. *For $G_t \sim \mathtt{DD}(t,p,r)$ asymptotically as $s, t \to \infty$ we have*

$$\mathbb{E}[\deg_t(s)] = \begin{cases}\Theta(\log\frac{t}{s}) & \text{if } p = 0 \text{ and } r > 0\\ \Theta\left(\frac{t^p}{s^p}\right) & \text{if } 0 < p < \frac{1}{2} \text{ and } r > 0,\\ \Theta\left(\frac{t^p}{s^p}\log s\right) & \text{if } p = \frac{1}{2} \text{ and } r > 0,\\ \Theta\left(\frac{t^p}{s^p}s^{2p-1}\right) & \text{if } p > \frac{1}{2} \text{ or } r = 0.\end{cases}$$

Let us note that when $s = \Theta(t)$, the asymptotics of $\mathbb{E}[\deg_t(s)]$ are exactly like those for $\mathbb{E}[\deg_t(t)]$ and $\mathbb{E}[D(G_t)]$, only the leading coefficients are different for each case. For different ranges of p and r, we have rates of growth equal to $\Theta(1)$, $\Theta(\log t)$ or $\Theta(t^{2p-1})$, respectively.

Finally, we present bounds for the deviation of $\deg_t(s)$ from its mean when $s = O(1)$.

Theorem 6. *Asymptotically for $G_t \sim \mathtt{DD}(t,p,r)$ and $s = O(1)$ it holds that*

$$\Pr[\deg_t(s) \geq A\,C\,t^p \log^2(t)] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 0$.

Theorem 7. *For $G_t \sim \mathtt{DD}(t,p,r)$ with $p > 0$ and $s = O(1)$ it holds that*

$$\Pr\left[\deg_t(s) \leq \frac{C}{A} t^p \log^{-3-\varepsilon}(t)\right] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 0$.

The proofs are analogous to those for the tails of the distribution of $D(G_t)$, so we omit sketches and refer the reader directly to Appendices F and G.

3 Discussion

In this paper we have focused on rigorous and precise analyses of the average degree $D(G_t)$ and and a fixed given node degree $\deg_t(s)$ in the divergence-duplication graph. We have derived asymptotic expressions for the expected values of these quantities and have also shown that with high probability they are only polylogarithmic factors away from the means of their expected values.

It is worth pointing that it is the parameter p that drives the rate of growth of the expected value for these parameters. We note that exact analysis reveals the fact that the value of parameter r and the structure of the starting graph G_{t_0} impact only the leading constants and lower order terms.

We observe that there are several phase transitions of these quantities as a function of p and r. This distinguishes it from the preferential graph model [9]. However, as demonstrated in [13], it seems that all real-world networks fall within a range $\frac{1}{2} < p < 1$, $r > 0$ – and this case should probably be the main topic of further investigation.

Future work may go along the lines of investigating further properties of the degree distribution as a function of both degree and time t. For example we might investigate the number of nodes of (given) degree k or the maximum degree in the graph. The latter is clearly bounded from below by $\deg_t(1)$, so from Theorem 7 one concludes that it is a polylogarithmic factor below t^p – but the upper bound still remains an open question since it requires bounds on the right tail of $\deg_t(t)$. The problem is that $\deg_t(t)$ is depends on the whole degree distribution in G_{t-1}, and therefore it is very unlikely that for s closer to t we have similar tail bounds as for $\deg_t(s)$ when $s = O(1)$.

In order to study graph symmetry we need more information about the degree distribution. This will allow us to find the ranges of parameters (p, r) for which we obtain an asymmetric graph with high probability or the ranges where non-negligible symmetries occur. In other words, we could explain theoretically Figure 1. This in turn will lead to finding efficient algorithms for graph compression extending [2, 9] to the duplication-divergence model. Moreover, the degree distribution may be useful in the problem of inferring node arrival order in networks [10].

References

1. Abramowitz, M., Stegun, I.: Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55. Dover Publications (1972)
2. Choi, Y., Szpankowski, W.: Compression of graphical structures: Fundamental limits, algorithms, and experiments. IEEE Transactions on Information Theory **58**(2), 620–638 (2012)
3. Colak, R., Hormozdiari, F., Moser, F., Schönhuth, A., Holman, J., Ester, M., Sahinalp, S.C.: Dense graphlet statistics of protein interaction and random networks. In: Biocomputing 2009, pp. 178–189. World Scientific Publishing, Singapore (2009)
4. Diestel, R.: Graph Theory. Springer (2005)
5. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. ACM SIGCOMM Computer Communication Review **29**(4), 251–262 (1999)
6. Hermann, F., Pfaffelhuber, P.: Large-scale behavior of the partial duplication random graph. ALEA **13**, 687–710 (2016)
7. Jordan, J.: The connected component of the partial duplication graph. ALEA – Latin American Journal of Probability and Mathematical Statistics **15**, 1431–1445 (2018)
8. Li, S., Choi, K.P., Wu, T.: Degree distribution of large networks generated by the partial duplication model. Theoretical Computer Science **476**, 94–108 (2013)
9. Łuczak, T., Magner, A., Szpankowski, W.: Asymmetry and structural information in preferential attachment graphs. Random Structures and Algorithms **55**(3), 696–718 (2019)
10. Magner, A., Sreedharan, J., Grama, A., Szpankowski, W.: Inferring temporal information from a snapshot of a dynamic network. Nature Scientific Reports **9**, 3057–3062 (2019)
11. Shao, M., Yang, Y., Guan, J., Zhou, S.: Choosing appropriate models for protein–protein interaction networks: a comparison study. Briefings in Bioinformatics **15**(5), 823–838 (2013)
12. Solé, R., Pastor-Satorras, R., Smith, E., Kepler, T.: A model of large-scale proteome evolution. Advances in Complex Systems **5**(01), 43–54 (2002)
13. Sreedharan, J., Turowski, K., Szpankowski, W.: Revisiting Parameter Estimation in Biological Networks: Influence of Symmetries (2019)
14. Szpankowski, W.: Average case analysis of algorithms on sequences. John Wiley & Sons (2011)
15. Turowski, K., Magner, A., Szpankowski, W.: Compression of Dynamic Graphs Generated by a Duplication Model. In: 56th Annual Allerton Conference on Communication, Control, and Computing. pp. 1089–1096 (2018)
16. Turowski, K., Sreedharan, J., Szpankowski, W.: Temporal Ordered Clustering in Dynamic Networks (2019)
17. Zhang, J.: Evolution by gene duplication: an update. Trends in Ecology & Evolution **18**(6), 292–298 (2003)

Appendix

A Proof of Lemma 1

We first observe that it follows from the definition of the model that the degree of the new vertex $t+1$ is the total number of edges from $t+1$ to $N_t(parent(t+1))$ (chosen independently with probability p) and to all other vertices (chosen independently with probability $\frac{r}{t}$). Note that it can be expressed as a sum of two independent binomial variables

$$\deg_{t+1}(t+1) \sim \mathrm{Bin}\left(\deg_t(parent(t+1)), p\right) + \mathrm{Bin}\left(t - \deg_t(parent(t+1)), \frac{r}{t}\right).$$

Hence

$$\begin{aligned}
\mathbb{E}[\deg_{t+1}(t+1) \mid G_t] &= \sum_{k=0}^{t} \Pr(\deg_t(parent(t+1)) = k) \sum_{a=0}^{k} \binom{k}{a} p^a (1-p)^{k-a} \\
&\quad \sum_{b=0}^{t-k} \binom{t-k}{b} \left(\frac{r}{t}\right)^b \left(1-\frac{r}{t}\right)^{t-k-b} (a+b) \\
&= \sum_{k=0}^{t} \Pr(\deg_t(parent(t+1)) = k) \left(pk + \frac{r}{t}(t-k)\right) \\
&= \left(p - \frac{r}{t}\right) \sum_{k=0}^{t} k \Pr(\deg_t(parent(t+1)) = k) + r.
\end{aligned}$$

Since parent sampling is uniform, we know that $\Pr(parent(t+1) = i) = \frac{1}{t}$ and therefore

$$D(G_t) = \sum_{i=1}^{t} \Pr(parent(t+1) = i) \deg_t(i) = \sum_{k=0}^{t} k \Pr(\deg_t(parent(t+1)) = k).$$

Combining the last two equations above with the law of total expectation we finally establish Lemma 1.

B Proof of Lemma 2

Let the process $(M_n)_{n=n_0}^{\infty}$ be defined as $M_{n_0} = f(G_{n_0})$ and

$$M_n = f(G_n) \prod_{k=n_0}^{n-1} \frac{1}{g_1(k)} - \sum_{j=n_0}^{n-1} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)}$$

Observe that

$$\mathbb{E}[M_{n+1} \mid G_n] = \mathbb{E}[f(G_{n+1}) \mid G_n] \prod_{k=n_0}^{n} \frac{1}{g_1(k)} - \sum_{j=n_0}^{n} g_2(j) \prod_{k=n_0}^{j} \frac{1}{g_1(k)}$$

$$= f(G_n)\prod_{k=n_0}^{n-1}\frac{1}{g_1(k)} - \sum_{j=n_0}^{n-1} g_2(j)\prod_{k=n_0}^{j}\frac{1}{g_1(k)} = M_n$$

which proves that $(M_n)_{n=n_0}^{\infty}$ is a martingale.

Furthermore, after some algebra and taking expectation with respect to G_n we arrive at

$$\begin{aligned}\mathbb{E}f(G_n) &= \mathbb{E}[M_n]\prod_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j)\prod_{k=n_0}^{j}\frac{1}{g_1(k)}\prod_{k=n_0}^{n-1} g_1(k) \\ &= f(G_{n_0})\prod_{k=n_0}^{n-1} g_1(k) + \sum_{j=n_0}^{n-1} g_2(j)\prod_{k=j+1}^{n-1} g_1(k)\end{aligned}$$

which completes the proof.

C Proof of Lemma 5

We estimate the sum using Lemma 4 and the Euler-Maclaurin formula [14, p. 294]. For the first case we have

$$\begin{aligned}&\sum_{j=n_0}^{n}\frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} = \sum_{j=n_0}^{n} j^{a-b}\left(1+O\left(\frac{1}{j}\right)\right) = \int_{n_0}^{n} j^{a-b}\left(1+O\left(\frac{1}{j}\right)\right)\mathrm{d}j \\ &= \left[j^{a-b+1}\left(\frac{1}{a-b+1}+O\left(\frac{1}{j}\right)\right)\right]_{n_0}^{n} = n^{a-b+1}\left(\frac{1}{a-b+1}+O\left(\frac{1}{n}\right)\right)+O(1).\end{aligned}$$

The second case is similar

$$\sum_{j=n_0}^{n}\frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} = \int_{n_0}^{n}\frac{1}{j}\left(1+O\left(\frac{1}{j}\right)\right)\mathrm{d}j = \ln n + O(1).$$

For the third case we first find that

$$\begin{aligned}&\sum_{j=n}^{\infty}\frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} = \sum_{j=n}^{\infty} j^{a-b}\left(1+O\left(\frac{1}{j}\right)\right) = \int_{n}^{\infty} j^{a-b}\left(1+O\left(\frac{1}{j}\right)\right)\mathrm{d}j \\ &\quad = \left[j^{a-b+1}\left(\frac{1}{a-b+1}+O\left(\frac{1}{j}\right)\right)\right]_{n}^{\infty} = n^{a-b+1}\left(\frac{1}{b-a-1}+O\left(\frac{1}{n}\right)\right).\end{aligned}$$

Finally, by the identical reasoning it is true that

$$\sum_{j=n}^{\infty}\frac{\prod_{i=1}^{k}\Gamma(j+a_i)}{\prod_{i=1}^{k}\Gamma(j+b_i)} = O(1),$$

and subtracting one from another completes the proof.

D Proof of Theorem 2

In order to prove the theorem we proceed as following: first we provide an asymptotic bound on $\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right]$, then we apply it for a suitable choices of λ, which allow us to use Chernoff bound.

Lemma 6. *For any $\lambda = O(\frac{1}{t})$ it holds that*

$$\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \leq \exp\left(\lambda p D(G_t)(1+O(\lambda t)) + \lambda r(1+O(\lambda))\right).$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \\
&= \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}\left[\exp\left(\lambda Bin(\deg_t(i), p) + \lambda Bin\left(t - \deg_t(i), \frac{r}{t}\right)\right)|G_t\right] \\
&\leq \frac{1}{t}\sum_{i=1}^{t} \left(1 - p + pe^{\lambda}\right)^{\deg_t(i)} \left(1 - \frac{r}{t} + \frac{r}{t}e^{\lambda}\right)^{t-\deg_t(i)}.
\end{aligned}$$

Since $e^x \leq 1+x+x^2$ for all $x \in [0,1]$, $(1+x)^y \leq 1+xy+(xy)^2$ for $0 \leq xy \leq 1$ and $1+x \leq e^x$ for any x:

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \\
&\leq \frac{1}{t}\sum_{i=1}^{t} (1 + p\lambda(1+O(\lambda))^{\deg_t(i)} \left(1 + \frac{r\lambda}{t}(1+O(\lambda))\right)^{t-\deg_t(i)} \\
&\leq \frac{1}{t}\sum_{i=1}^{t} (1 + p\lambda \deg_t(i)(1+O(\lambda t)) \left(1 + r\lambda(1+O(\lambda)\right)) \\
&\leq \frac{1}{t}\sum_{i=1}^{t} (1 + p\lambda \deg_t(i)(1+O(\lambda t))) \exp\left(r\lambda(1+O(\lambda))\right) \\
&= (1 + p\lambda D(G_t)(1+O(\lambda t))) \exp\left(r\lambda(1+O(\lambda))\right) \\
&\leq \exp\left(\lambda p D(G_t)(1+O(\lambda t)) + \lambda r(1+O(\lambda))\right).
\end{aligned}$$

Now we are ready to finally prove the theorem.

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_{t+1} D(G_{t+1})\right) \mid G_t\right] \\
&= \mathbb{E}\left[\exp\left(\lambda_{t+1}\left(\frac{t}{t+1}D(G_t) + \frac{2}{t+1}\deg_{t+1}(t+1)\right)\right) \mid G_t\right] \\
&= \exp\left(\frac{\lambda_{t+1} t}{t+1} D(G_t)\right) \mathbb{E}\left[\exp\left(\frac{2\lambda_{t+1}}{t+1}\deg_{t+1}(t+1)\right) \mid G_t\right]
\end{aligned}$$

Now we may use Lemma 7 with $\lambda = \frac{2\lambda_{t+1}}{t+1}$ to get

$$\mathbb{E}\left[\exp\left(\lambda_{t+1} D(G_{t+1})\right) \mid G_t\right] =$$

$$\le \exp\left(\lambda_{t+1} D(G_t)\left(1-\frac{2p-1}{t+1}\right)(1+O(\lambda_{t+1}))+\frac{2r\lambda_{t+1}}{t+1}(1+o(t^{-1}))\right).$$

Let us define for $k = t_0, \ldots, t-1$

$$\lambda_k = \lambda_{k+1}\left(1+\left(\frac{2p-1}{t+1}\right)(1+O(\lambda_{k+1}))\right)$$

and let $\varepsilon_t \ge \lambda_k$ for all k.

Then clearly

$$\begin{aligned}\lambda_{t_0} &\in \left[\lambda_t \prod_{k=t_0}^{t-1}\left(1+\frac{2p-1}{k+1}\right), \lambda_t \prod_{k=t_0}^{t-1}\left(1+\left(\frac{2p-1}{k+1}\right)(1+O(\varepsilon_t))\right)\right] \\ &\subseteq \left[\lambda_t\left(\frac{t}{t_0}\right)^{2p-1}(1+o(1)), \lambda_t\left(\frac{t}{t_0}\right)^{(2p-1)(1+O(\varepsilon_t))}(1+o(1))\right]\end{aligned}$$

It follows that

$$\mathbb{E}\left[\exp\left(\lambda_t D(G_t)\right)\right] \le \exp\left(\lambda_{t_0} D(G_{t_0})\right)\prod_{k=t_0}^{t-1}\exp\left(\frac{2r\lambda_{k+1}}{k+1}\left(1+o(k^{-1})\right)\right) \tag{3}$$

$$\le \exp\left(\lambda_{t_0} D(G_{t_0})\right)\exp\left(2r\varepsilon_{t+1}\ln\frac{t}{t_0}+C_1\right) = \exp\left(\lambda_{t_0} D(G_{t_0})\right)\left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1}+C_1} \tag{4}$$

for a certain constant C_1.

Finally, let $\lambda_t = \varepsilon_t\left(\frac{t}{t_0}\right)^{-(2p-1)(1+O(\varepsilon_t)))}$ so that $\lambda_{t_0} \le \varepsilon_t$. Then from Chernoff bound it follows that

$$\begin{aligned}\Pr[D(G_t) \ge \alpha\mathbb{E}D(G_t)] &= \Pr[\exp(D(G_t) - \alpha\mathbb{E}D(G_t)) \ge 1] \\ &\le \exp\left(-\alpha\lambda_t\mathbb{E}D(G_t)\right)\mathbb{E}[\exp\left(\lambda_t D(G_t)\right)] \\ &\le \exp\left(-\alpha\lambda_t\mathbb{E}D(G_t)\right)\exp\left(\lambda_{t_0} D(G_{t_0})\right)\left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1}+C_1}\end{aligned}$$

Assume $\varepsilon_t = \frac{1}{\ln(t/t_0)}$. For $p > \frac{1}{2}$ we have $\mathbb{E}D(G_t) = C_2\left(\frac{t}{t_0}\right)^{2p-1}(1+o(1))$, and therefore

$$\begin{aligned}&\Pr\left[D(G_t) \ge \alpha C_2\left(\frac{t}{t_0}\right)^{2p-1}(1+o(1))\right] \\ &\le \exp\left(-\alpha C_2\varepsilon_t\left(\frac{t}{t_0}\right)^{-(2p-1)\varepsilon_t}\right)\exp\left(\varepsilon_t(t_0-1))\right)\left(\frac{t}{t_0}\right)^{2r\varepsilon_{t+1}+C_1} \\ &\le \exp\left(-\alpha C_2\frac{\exp\left(-2p+1\right)}{\ln\left(t/t_0\right)}\right)\exp\left(\frac{t_0-1}{\ln\left(t/t_0\right)}\right)\exp\left(2r+C_1\right)\end{aligned}$$

The last two elements are bounded by a constant, so it is sufficient to pick $\alpha = \frac{A}{C_2}\exp(2p-1)\ln^2(t)$ to complete the proof for the case $p > \frac{1}{2}$.

Now, for $p < \frac{1}{2}$ and $p = \frac{1}{2}$ it is sufficient to use $\mathbb{E}D(G_t) = C_2(1+o(1))$ and $\mathbb{E}D(G_t) = C_2 \ln t(1+o(1))$, respectively.

E Proof of Theorem 3

We start the proof by obtaining a simple lemma, analogous to Lemma 6:

Lemma 7. *For any $\lambda = O(\frac{1}{t})$ it holds that*

$$\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \le \exp\left(2\lambda p D(G_t)(1+O(\lambda)) + 2\lambda r(1+O(\lambda))\right).$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \\
&\quad = \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}\left[\exp\left(\lambda Bin(\deg_t(i), p) + \lambda Bin\left(t - \deg_t(i), \frac{r}{t}\right)\right)|G_t\right] \\
&\quad \le \frac{1}{t}\sum_{i=1}^{t}\left(1 - p + pe^{\lambda}\right)^{\deg_t(i)}\left(1 - \frac{r}{t} + \frac{r}{t}e^{\lambda}\right)^{t-\deg_t(i)}.
\end{aligned}$$

Since $e^x \le 1 + x + x^2$ for all $x \in [0,1]$, $(1+x)^y \le 1 + 2xy$ for $0 \le xy \le 1$, and $1 + x \le e^x$ for all x

$$\begin{aligned}
&\mathbb{E}\left[\exp(\lambda \deg_{t+1}(t+1))|G_t\right] \\
&\quad \le \frac{1}{t}\sum_{i=1}^{t}(1 + p\lambda(1+O(\lambda))^{\deg_t(i)}\left(1 + \frac{r\lambda}{t}(1+O(\lambda))\right)^{t-\deg_t(i)} \\
&\quad \le \frac{1}{t}\sum_{i=1}^{t}(1 + 2p\lambda\deg_t(i)(1+O(\lambda)))\,(1 + 2r\lambda(1+O(\lambda)))) \\
&\quad \le \frac{1}{t}\sum_{i=1}^{t}(1 + 2p\lambda\deg_t(i)(1+O(\lambda)))\exp\left(2r(1+O(\lambda))\right) \\
&\quad = (1 + 2p\lambda D(G_t)(1+O(\lambda)))\exp\left(2r(1+O(\lambda)))\right) \\
&\quad \le \exp\left(2\lambda p D(G_t)(1+O(\lambda)) + 2\lambda r(1+O(\lambda))\right).
\end{aligned}$$

Next, using the lemma above and Theorem 2 we limit the growth of $D(G_t)$ over certain intervals:

Lemma 8. *Let $p > \frac{1}{2}$. For sufficiently large t and all $k < t$ it is true that*

$$\Pr[D(G_{(k+1)t}) - D(G_{kt}) \ge AC((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 1$.

Proof. First, let us define events $\mathcal{B}_i = [D(G_{i+1}) \geq (A+1)\, C_1\, i^{2p-1} \log^2(i)]$ with a constant C_1 such that by Theorem 2 it is true that $\Pr[\mathcal{B}_i] = O(i^{-A-1})$. Let us also denote $\mathcal{A}_k = \bigcup_{i=kt}^{(k+1)t-1} \mathcal{B}_i$ and observe that $\Pr[\mathcal{A}_k] = O(t^{-A})$.

Now, we note that from Lemma 6 for any $\lambda = o(1)$

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda(D(G_{t+1}) - D(G_t))\right) \middle| G_t, \neg\mathcal{B}_t\right] \\
&\quad\leq \mathbb{E}\left[\exp\left(\frac{2\lambda}{t+1} \deg_{t+1}(t+1)\right) \middle| G_t, \neg\mathcal{B}_t\right] \\
&\quad\leq \left[\exp\left(\frac{2\lambda p}{t+1} D(G_t)(1+O(\lambda)) + \frac{2\lambda r}{t+1}(1+O(\lambda))\right) \middle| \neg\mathcal{B}_t\right] \\
&\quad\leq \exp\left(\lambda\,(A+1)\, C_2\, t^{2p-2} \log^2(t)(1+o(1))\right)
\end{aligned}$$

for a certain constant C_2.

Now we proceed as following:

$$\begin{aligned}
&\Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq d | G_{kt}] \\
&\quad\leq \Pr[D(G_{(k+1)t}) - D(G_{kt}) \geq d | G_{kt}, \neg\mathcal{A}_k] \Pr[\neg\mathcal{A}] + \Pr[\mathcal{A}_k] \\
&\quad\leq \exp(-\lambda d)\mathbb{E}\left[\exp\left(\lambda(D(G_{(k+1)t}) - D(G_{kt}))\right) | G_{kt}, \neg\mathcal{A}_k\right] + O(t^{-A}) \\
&\quad\leq \exp(-\lambda d) \prod_{i=kt}^{(k+1)t-1} \mathbb{E}\left[\exp\left(\lambda(D(G_{i+1}) - D(G_i))\right) \middle| G_i, \neg\mathcal{B}_i\right] + O(t^{-A}) \\
&\quad\leq \exp(-\lambda d) \prod_{i=kt}^{(k+1)t-1} \exp\left(\lambda\,(A+1)\, C_2\, i^{2p-2} \log^2(i)(1+o(1))\right) + O(t^{-A}) \\
&\quad\leq \exp(-\lambda d) \exp\left(\sum_{i=kt}^{(k+1)t-1} \lambda\,(A+1)\, C_3\, i^{2p-2} \log^2(t)(1+o(1))\right) + O(t^{-A}) \\
&\quad\leq \exp(-\lambda d) \exp\left(\lambda\,(A+1)\, C_3((k+1)^{2p-1} - k^{2p-1}) t^{2p-1} \log^2(t)\right) + O(t^{-A})
\end{aligned}$$

for a certain constant C_3.

Finally, it is sufficient to take $\lambda = \left(((k+1)^{2p-1} - k^{2p-1}) \log^2(t)\right)^{-1}$ and $d = AC_4((k+1)^{2p-1} - k^{2p-1}) t^{2p-1} \log^2(t)$ for sufficiently large C_4 to obtain the final result.

Now we may return to the main theorem. Let $Y_k = D(G_{(k+1)t}) - D(G_{kt})$. We know that for $p > \frac{1}{2}$

$$\mathbb{E}Y_k = \mathbb{E}D(G_{(k+1)t}) - \mathbb{E}D(G_{kt}) = C_1 \left((k+1)^{2p-1} - k^{2p-1}\right) t^{2p-1}(1+o(1))$$

for some constant C_1.

Let now define the following events:

$$\mathcal{A}_1 = \left[Y_k \leq \frac{t^{2p-1}}{f(t)}\right]$$

$$\mathcal{A}_2 = \left[\frac{t^{2p-1}}{f(t)} < Y_k \leq C_2((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)\right]$$
$$\mathcal{A}_3 = \left[Y_k > C_2((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t)\right]$$

for a constant C_2 such that (from the lemma above) $\Pr[\mathcal{A}_3] = O(t^{-2})$. Here $f(t)$ is any (monotonic) function such that $f(t) \to \infty$ as $t \to \infty$.

We know that

$$\begin{aligned}
\mathbb{E}Y_k &= \mathbb{E}\left[Y_k|\mathcal{A}_1\right]\Pr\left[\mathcal{A}_1\right] + \mathbb{E}\left[Y_k|\mathcal{A}_2\right]\Pr\left[\mathcal{A}_2\right] + \mathbb{E}\left[Y_k|\mathcal{A}_3\right]\Pr\left[\mathcal{A}_3\right] \\
\mathbb{E}Y_k &\geq C_1\left((k+1)^{2p-1} - k^{2p-1}\right)t^{2p-1} \\
\mathbb{E}\left[Y_k|\mathcal{A}_1\right] &\leq \frac{t^{2p-1}}{f(t)} \\
\mathbb{E}\left[Y_k|\mathcal{A}_2\right] &\leq C_2((k+1)^{2p-1} - k^{2p-1})t^{2p-1}\log^2(t) \\
\mathbb{E}\left[Y_k|\mathcal{A}_3\right] &\leq (k+1)t
\end{aligned}$$

and therefore for sufficiently large t it holds that

$$\begin{aligned}
\Pr[\mathcal{A}_1] &\leq \frac{C_2\left((k+1)^{2p-1} - k^{2p-1}\right)\log^2(t) - C_1\left((k+1)^{2p-1} - k^{2p-1}\right)}{C_2\left((k+1)^{2p-1} - k^{2p-1}\right)\log^2(t) - \frac{1}{f(t)}} \\
&\leq 1 - \frac{C_1}{2C_2\log^2(t)}.
\end{aligned}$$

Let now $\tau = kt$.

$$\begin{aligned}
\Pr\left[D(G_\tau) \leq t^{2p-1}f^{-1}(t)\right] &= \Pr\left[\bigcap_{i=1}^{k} Y_i \leq \frac{t^{2p-1}}{f(t)}\right] \\
&\leq \prod_{i=1}^{k}\Pr\left[Y_i \leq \frac{t^{2p-1}}{f(t)}\right] \leq \prod_{i=1}^{k}\left(1 - \frac{C_1}{2C_2\log^2(t)}\right)
\end{aligned}$$

Therefore, if we assume $k = \frac{2AC_2}{C_1}\log^3(t)$, we get

$$\Pr\left[D(G_\tau) \leq \frac{t^{2p-1}}{f(t)}\right] = \exp\left(-A\log(t)\right) = O(t^{-A})$$

and finally

$$\Pr\left[D(G_t) \leq \frac{C_3}{A^{2p-1}}t^{2p-1}\log^{-3(2p-1)-\varepsilon}(t)\right] = O(t^{-A}).$$

for some constant C_3 and any $\varepsilon > 0$.

F Proof of Theorem 6

$$\mathbb{E}\left[\exp\left(\lambda_{t+1}\deg_{t+1}(s)\right) \mid G_t\right] =$$

$$
\begin{aligned}
&= \left(\frac{\deg_t(s)}{t} p + \frac{t-\deg_t(s)}{t}\frac{r}{t}\right)\exp\left(\lambda_{t+1}\left(\deg_t(s)+1\right)\right) \\
&\quad + \left(\frac{\deg_t(s)}{t}(1-p) + \frac{t-\deg_t(s)}{t}\left(1-\frac{r}{t}\right)\right)\exp\left(\lambda_{t+1}\deg_t(s)\right) \\
&= \exp\left(\lambda_{t+1}\deg_t(s)\right) \\
&\quad \left(\frac{\deg_t(s)}{t}\left(1-p+p\exp\left(\lambda_{t+1}\right)\right) + \frac{t-\deg_t(s)}{t}\left(1-\frac{r}{t}+\frac{r}{t}\exp\left(\lambda_{t+1}\right)\right)\right) \\
&\le \exp\left(\lambda_{t+1}\deg_t(s)\right)\left(1+\left(\frac{p\deg_t(s)}{t}+\frac{r\left(t-\deg_t(s)\right)}{t^2}\right)\left(\lambda_{t+1}+\lambda_{t+1}^2\right)\right) \\
&\le \exp\left(\lambda_{t+1}\deg_t(s)+\left(\frac{p\deg_t(s)}{t}+\frac{r\left(t-\deg_t(s)\right)}{t^2}\right)\left(\lambda_{t+1}+\lambda_{t+1}^2\right)\right) \\
&= \exp\left(\lambda_{t+1}\deg_t(s)\left(1+\left(\frac{p}{t}-\frac{r}{t^2}\right)\left(1+\lambda_{t+1}\right)\right)\right)\exp\left(\lambda_{t+1}\left(1+\lambda_{t+1}\right)\frac{r}{t}\right).
\end{aligned}
$$

Let us assume that $\lambda_k \le \varepsilon_t = o(1)$ for all $s \le k \le t$. Then for all $k = s, s+1, \ldots, t$ we have

$$
\lambda_k = \lambda_{k+1}\left(1+\left(\frac{p}{k}-\frac{r}{k^2}\right)\left(1+\lambda_{k+1}\right)\right) \le \lambda_{k+1}\left(1+\left(\frac{p}{k}-\frac{r}{k^2}\right)\left(1+\varepsilon_t\right)\right)
$$

which lead us to

$$
\begin{aligned}
\lambda_s &\le \lambda_t \prod_{k=s}^{t-1}\left(1+\left(\frac{p}{k}-\frac{r}{k^2}\right)\left(1+\varepsilon_t\right)\right) \le \lambda_t \exp\left(\left(1+\varepsilon_t\right)\sum_{k=s}^{t-1}\left(\frac{p}{k}-\frac{r}{k^2}\right)\right) \\
&\le \lambda_t \exp\left(\left(1+\varepsilon_t\right)\int_s^t\left(\frac{p}{k}-\frac{r}{k^2}\,\mathrm{d}k\right)\right) \le \lambda_t\left(\frac{t}{s}\right)^{p(1+\varepsilon_t)}\exp\left(\frac{r}{t}\left(1+\varepsilon_t\right)\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda_t\deg_t(s)\right)|G_s\right] \le \exp\left(\lambda_s\deg_s(s)\right)\right)\prod_{k=s}^{t-1}\exp\left(\lambda_{k+1}\left(1+\lambda_{k+1}\right)\frac{r}{k}\right) \\
&\le \exp\left(\lambda_s\deg_s(s)\right)\right)\exp\left(\varepsilon_t\left(1+\varepsilon_t\right)r\ln\frac{t}{s}\right) \le \exp\left(\lambda_s\deg_s(s)\right)\right)\left(\frac{t}{s}\right)^{r\varepsilon_t(1+\varepsilon_t)}
\end{aligned}
$$

Now, let $\lambda_t = \epsilon_t\left(\frac{t}{s}\right)^{-p(1+\varepsilon_t)}\exp\left(-\frac{r}{t}\left(1+\varepsilon_t\right)\right)$ so that $\lambda_s \le \epsilon_t$. Then, from Chernoff bound it follows that

$$
\begin{aligned}
\Pr[\deg_t(s) \ge \alpha\mathbb{E}\deg_t(s)|G_s] &= \Pr[\exp(\deg_t(s)-\alpha\mathbb{E}\deg_t(s)) \ge 1|G_s] \\
&\le \exp\left(-\alpha\lambda_t\mathbb{E}[\deg_t(s)|G_s]\right)\mathbb{E}[\exp\left(\lambda_t\deg_t(s)\right)|G_s] \\
&\le \exp\left(-\alpha\lambda_t\mathbb{E}[\deg_t(s)|G_s]\right)\exp\left(\lambda_s\deg_s(s)\right)\left(\frac{t}{s}\right)^{r\varepsilon_t(1+\varepsilon_t)}.
\end{aligned}
$$

Let's assume $\varepsilon_t = \frac{1}{\ln t}$. From Theorem 4 we know that if $s = O(1)$, then it holds that $\mathbb{E}[\deg_t(s)|G_s] = C_1 t^p$ and therefore

$$
\Pr[\deg_t(s) \ge \alpha C_1 t^p|G_s] \le \exp\left(-\alpha C_2\epsilon_t t^{-p\varepsilon_t}\right)\exp\left(\epsilon_t\deg_s(s)\right)\right)\left(\frac{t}{s}\right)^{r\varepsilon_t(1+\varepsilon_t)}
$$

$$\leq \exp\left(-\frac{\alpha C_3}{\ln t}\right)\exp\left(\frac{\deg_s(s)}{\ln t}\right)\exp(2r)$$

for certain constants C_2, C_3.

Therefore, it is sufficient to set $\alpha = \frac{A}{C_3}\ln^2 t$ to get the final result.

G Proof of Theorem 7

We proceed similarly as in the proof of Theorem 3:

Lemma 9. *Let $p > 0$ and $s = O(1)$. For sufficiently large t and all $k < t$ it is true that*

$$\Pr[\deg_{(k+1)t}(s) - \deg_{kt}(s) \geq AC((k+1)^p - k^p)t^p \log^2(t)] = O(t^{-A})$$

for some fixed constant $C > 0$ and any $A > 1$.

Proof. Let us define events $\mathcal{B}_i = [\deg_{i+1}(s) \geq (A+1)\, C_1\, i^p \log^2(i)]$ with a constant C_1 such that by Theorem 6 it is true that $\Pr[\mathcal{B}_i] = O(i^{-A-1})$.

Now, for any $\lambda = o(1)$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\exp\big(\lambda(\deg_{t+1}(s) - \deg_t(s))\big)\,\middle|\, G_t, \neg\mathcal{B}_t\right] \\
&= \left[\frac{\deg_t(s)}{t}(1 - p + p\exp(\lambda)) + \frac{t - \deg_t(s)}{t}\left(1 - \frac{r}{t} + \frac{r}{t}\exp(\lambda)\right)\,\middle|\, \neg\mathcal{B}_t\right] \\
&\leq \exp\left(\left(\frac{p\deg_t(s)}{t} + \frac{r(t - \deg_t(s))}{t^2}\right)(\lambda + \lambda^2)\right) \\
&\leq \exp\big(\lambda(A+1)\, C_1\, pt^{p-1}\log^2(t)\,(1 + o(1))\big)\,.
\end{aligned}$$

Let us now denote $\mathcal{A}_k = \bigcup_{i=kt}^{(k+1)t-1} \mathcal{B}_i$ and observe that $\Pr[\mathcal{A}_k] = O(t^{-A})$. We proceed similarly to the proof of Theorem 3:

$$\begin{aligned}
&\Pr[\deg_{(k+1)t}(s) - \deg_{kt}(s) \geq d | G_{kt}] \\
&\leq \Pr[\deg_{(k+1)t}(s) - \deg_{kt}(s) \geq d | G_{kt}, \neg\mathcal{A}_k]\Pr[\neg\mathcal{A}] + \Pr[\mathcal{A}_k] \\
&\leq \exp(-\lambda d)\mathbb{E}\left[\exp\Big(\lambda(\deg_{(k+1)t}(s) - \deg_{kt}(s))\Big) | G_{kt}, \neg\mathcal{A}_k\right] + O(t^{-A}) \\
&\leq \exp(-\lambda d)\prod_{i=kt}^{(k+1)t-1}\mathbb{E}\left[\exp\big(\lambda(\deg_{i+1}(s) - \deg_i(s))\big)\,\middle|\, G_i, \neg\mathcal{B}_i\right] + O(t^{-A}) \\
&\leq \exp(-\lambda d)\prod_{i=kt}^{(k+1)t-1}\exp\big(\lambda\,(A+1)\, C_1\, i^{p-1}\log^2(i)(1 + o(1))\big) + O(t^{-A}) \\
&\leq \exp(-\lambda d)\exp\left(\sum_{i=kt}^{(k+1)t-1}\lambda\,(A+1)\, C_1\, i^{p-1}\log^2(t)(1 + o(1))\right) + O(t^{-A})
\end{aligned}$$

$$\leq \exp(-\lambda d) \exp\left(\lambda\,(A+1)\,C_2((k+1)^p - k^p)t^p \log^2(t)\right) + O(t^{-A})$$

for a certain constant C_2.

Therefore, it is sufficient to take $\lambda = \left(((k+1)^p - k^p)\log^2(t)\right)^{-1}$ and $d = AC_3((k+1)^p - k^p)t^p \log^2(t)$ for sufficiently large C_3 to obtain the final result.

Now we return to the proof of the main theorem. Let $Z_k = \deg_{(k+1)t}(s) - \deg_{kt}(s)$. We know that for $p > 0$

$$\mathbb{E}Z_k = \mathbb{E}D(G_{(k+1)t}) - \mathbb{E}D(G_{kt}) = C_1\left((k+1)^p - k^p\right)t^p(1+o(1))$$

for some constant C_1.

Let now define the following events:

$$\begin{aligned}
\mathcal{A}_1 &= \left[Z_k \leq \frac{t^p}{f(t)}\right] \\
\mathcal{A}_2 &= \left[\frac{t^p}{f(t)} < Z_k \leq C_2((k+1)^p - k^p)t^p \log^2(t)\right] \\
\mathcal{A}_3 &= \left[Z_k > C_2((k+1)^p - k^p)t^p \log^2(t)\right]
\end{aligned}$$

for a constant C_2 such that (from the lemma above) $\Pr[\mathcal{A}_3] = O(t^{-2})$. Here $f(t)$ is any (monotonic) function such that $f(t) \to \infty$ as $t \to \infty$.

We know that

$$\begin{aligned}
\mathbb{E}Z_k &= \mathbb{E}\left[Z_k|\mathcal{A}_1\right]\Pr\left[\mathcal{A}_1\right] + \mathbb{E}\left[Z_k|\mathcal{A}_2\right]\Pr\left[\mathcal{A}_2\right] + \mathbb{E}\left[Z_k|\mathcal{A}_3\right]\Pr\left[\mathcal{A}_3\right] \\
\mathbb{E}Z_k &\geq C_1\left((k+1)^p - k^p\right)t^{2p-1} \\
\mathbb{E}\left[Z_k|\mathcal{A}_1\right] &\leq \frac{t^{2p-1}}{f(t)} \\
\mathbb{E}\left[Z_k|\mathcal{A}_2\right] &\leq C_2((k+1)^p - k^p)t^p \log^2(t) \\
\mathbb{E}\left[Z_k|\mathcal{A}_3\right] &\leq (k+1)t
\end{aligned}$$

and therefore for sufficiently large t it holds that

$$\begin{aligned}
\Pr[\mathcal{A}_1] &\leq \frac{C_2\left((k+1)^p - k^p\right)\log^2(t) - C_1\left((k+1)^p - k^p\right)}{C_2\left((k+1)^p - k^p\right)\log^2(t) - \frac{1}{f(t)}} \\
&\leq 1 - \frac{C_1}{2C_2\log^2(t)}.
\end{aligned}$$

Let now $\tau = kt$. Then,

$$\Pr\left[D(G_\tau) \leq t^p f^{-1}(t)\right] = \Pr\left[\bigcap_{i=1}^{k} Y_i \leq \frac{t^p}{f(t)}\right] \leq \prod_{i=1}^{k}\left(1 - \frac{C_1}{2C_2\log^2(t)}\right).$$

Therefore, if we assume $k = \frac{2AC_2}{C_1}\log^3(t)$, we get

$$\Pr\left[D(G_\tau) \leq \frac{t^p}{f(t)}\right] = \exp\left(-A\log(t)\right) = O(t^{-A})$$

and finally

$$\Pr\left[D(G_t) \leq \frac{C_3}{A^p} t^p \log^{-3p-\varepsilon}(t)\right] = O(t^{-A}).$$

for some constant C_3 and any $\varepsilon > 0$.